

21世纪高等学校规划教材 | 计算机科学与技术

数据结构与算法

(第2版)

文益民 张瑞霞 李健 编著



清华大学出版社



数据结构与算法

(第2版)

文益民 张瑞霞 李健 编著

清华大学出版社
北京

内 容 简 介

本书以提高学生的程序设计能力为主旨,全面地介绍程序设计的基础知识、各种常用的数据结构以及排序、查找的各种算法及其应用。为了方便教学,各数据结构类型和基本运算首先用类C代码加以描述,并做了详细的注解。全书既注重原理,又强调实践,配有大量的图表和习题,概念讲解清楚,逻辑性强,可读性好。此外,本书的特点还有:首次尝试采用“任务驱动”方式来设计教学内容;书中有大量以“思考”形式出现的问题,能在恰当的时机激发思考,启发思维,方便应用于“翻转课堂”教学模式;使用脚注介绍计算科学发展史知识和其他相关知识,可拓展学生的知识范围。

本书可作为技术应用型本科院校计算机专业教材,也可作为参加全国计算机软件水平程序员级别考试提供参考,还可供广大从事计算机应用的科技人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据结构与算法/文益民,张瑞霞,李健编著.—2版.—北京:清华大学出版社,2017

(21世纪高等学校规划教材·计算机科学与技术)

ISBN 978-7-302-45369-7

I. ①数… II. ①文… ②张… ③李… III. ①数据结构—高等学校—教材 ②算法分析—高等学校—教材 IV. ①TP311.12

中国版本图书馆CIP数据核字(2016)第260828号

责任编辑:郑寅堃 王冰飞

封面设计:傅瑞学

责任校对:梁毅

责任印制:沈露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:北京嘉实印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:16.5

字 数:400千字

版 次:2008年2月第1版

2017年5月第2版

印 次:2017年5月第1次印刷

印 数:1~1500

定 价:39.00元

产品编号:061580-01

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和教学方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上。精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展,顺应并符合21世纪教学发展的规律,代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版

社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。推出的特色精品教材包括:

- (1) 21 世纪高等学校规划教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。
- (2) 21 世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。
- (3) 21 世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。
- (4) 21 世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。
- (5) 21 世纪高等学校规划教材·信息管理与信息系统。
- (6) 21 世纪高等学校规划教材·财经管理与应用。
- (7) 21 世纪高等学校规划教材·电子商务。
- (8) 21 世纪高等学校规划教材·物联网。

清华大学出版社经过三十多年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail: weijj@tup.tsinghua.edu.cn

前言

大数据时代已经到来,数据与算法是数据科学与工程的重要内容,而“数据结构”是计算机算法设计的重要理论和方法基础,它不仅是计算机类专业的核心课程,而且已成为其他专业的重要教学内容。“数据结构与算法”的教学目的是:让学生学会分析需要计算机处理的数据对象的特性,以选择适当的数据结构、存储结构从而选择相应的算法;初步掌握算法性能分析的方法;初步掌握将实际问题求解转化为算法,进而转化为计算机程序的能力;通过本课程的学习,使学生获得更进一步的程序设计技能训练,提高编程能力,进而提高计算机软件工程能力。

长久以来由于数据结构课程自身的抽象和严密,以及数据结构开设的时间通常在大一的第二学期,教师感觉数据结构难教,学生普遍反映数据结构难学,学生很难独立完成算法的实现。基于上述问题我们在编写本教材时充分考虑了学生的知识结构和教师的教学方法。本教材的编写遵循的原则是:既注重原理又注重实践;既注重抽象思维又注重形象思维;既方便自学又方便教学。本书的编写有以下特色。

(1) 采用“任务驱动”方式来设计教学内容。除第1章外,在每一章,必要的基础知识介绍后都安排了一个问题作为学习完本章后要解决的“任务”。该问题具有两个特征:①有一定的趣味性,能较好地激发学生的学习兴趣 and 解决问题的动力;②综合性较强,问题的解决需要使用到本章中的知识。

(2) 利用教材中的“思考”标志,提出问题拓展学生思维。“思考”不同于“习题”,“习题”的作用代替不了“思考”,因为“习题”在每个章节之后,一般要等到讲完一个章节才会遇到,因此“习题”对于课堂教学是滞后的。采用提示“思考”的方式可以在教学中恰到好处地启发学生的思维。教材使用“思考”标志通常有3种情况:一是提醒学生注意;二是启发学生基于当前知识基础进一步思考;三是提示本教材没有讲授的内容以引导学有余力的学生拓展自身知识面。另外,这些标志为“思考”的问题,可方便地应用于“翻转课堂”教学模式。

(3) 在计算机专业的核心基础课程中增加计算机科学文化的知识,使学生在学本教材的过程中,不但能学到专业知识,还能了解计算机科学发展历史的相关知识和数据结构课程与其他课程的联系;对提高学生学习本课程的兴趣,拓宽学生的知识面大有好处。

全书共分8章:第1章介绍数据结构和算法分析的基本概念及程序设计基础;第2~4章介绍线性结构及一部分与线性结构密切相关的非线性结构;第5章和第6章分别介绍树形结构和图结构;第7章和第8章分别介绍排序和查找。

本书可作为技术应用型本科院校计算机专业教材,也可为参加全国计算机软件水平程序员级别考试提供参考,还可供广大从事计算机应用的科技人员参考。本书讲授60课时左右,除第1章外每章可安排2课时上机实习。

本书是由文益民、张瑞霞、李健三位在多年从事计算机专业数据结构课程教学的经验基

基础上,经过多次反复磋商和共同讨论而定稿。全书由桂林电子科技大学文益民整体构思、统稿、修改,易新河、文博奚等为本书的编辑、排版做了很多工作。

由于编著者水平有限,书中难免存在不足和疏漏之处,殷切期望广大读者批评指正。

文 益 民

2017 年 1 月于桂林电子科技大学

图书资源支持

感谢您一直以来对清华版图书的支持和爱护。为了配合本书的使用,本书提供配套的素材,有需求的用户请到清华大学出版社主页(<http://www.tup.com.cn>)上查询和下载,也可以拨打电话或发送电子邮件咨询。

如果您在使用本书的过程中遇到了什么问题,或者有相关图书出版计划,也请您发邮件告诉我们,以便我们更好地为您服务。

我们的联系方式:

地 址: 北京海淀区双清路学研大厦 A 座 707

邮 编: 100084

电 话: 010-62770175-4604

资源下载: <http://www.tup.com.cn>

电子邮件: weijj@tup.tsinghua.edu.cn

QQ: 883604 (请写明您的单位和姓名)

用微信扫一扫右边的二维码,即可关注清华大学出版社公众号“书圈”。



扫一扫

资源下载、样书申请
新书推荐、技术交流

目 录

第 1 章 绪论	1
1.1 数据结构的基本概念	2
1.1.1 数据结构的实例	2
1.1.2 数据结构的概念	4
1.1.3 学习数据结构的理由	5
1.2 算法分析的基本概念	6
1.2.1 算法	6
1.2.2 算法效率的分析	7
1.2.3 算法效率的评价	7
1.3 程序设计基础	10
1.3.1 软件工程的基本概念	10
1.3.2 软件设计基础	12
1.3.3 编码基础	12
1.3.4 计算机体系结构基础	13
习题 1	14
第 2 章 线性表	18
2.1 线性表的基本概念	18
2.1.1 线性表的基本运算	18
2.1.2 一个有趣的问题	19
2.2 线性表的顺序表示	20
2.2.1 顺序表	20
2.2.2 顺序表的基本运算	21
2.2.3 顺序表的算法分析	24
2.3 线性表的链式表示	25
2.3.1 线性链表	25
2.3.2 线性链表的基本运算	26
2.3.3 顺序表和链式表的比较	30
2.4 双链表和循环链表	31
2.4.1 双链表	31
2.4.2 循环链表	33
2.5 线性表的应用	34

习题 2	37
第 3 章 栈和队列	40
3.1 栈的概念及运算.....	40
3.1.1 栈的概念	40
3.1.2 栈的基本运算	41
3.1.3 一个有趣的问题	41
3.2 栈的存储和实现.....	42
3.2.1 栈的顺序表示	42
3.2.2 栈的链式表示	45
3.3 栈的应用.....	47
3.3.1 数制转换	47
3.3.2 表达式求值	49
3.3.3 栈与递归	53
3.3.4 回溯法	57
3.4 队列的概念及基本运算.....	59
3.4.1 队列的概念	59
3.4.2 队列的基本运算	59
3.4.3 一个有趣的问题	60
3.5 队列的存储结构及运算.....	60
3.5.1 队列的顺序表示	60
3.5.2 循环队列	62
3.5.3 队列的链式表示	65
3.6 队列的应用.....	67
习题 3	71
第 4 章 串、广义表及数组.....	75
4.1 串的定义和基本运算.....	75
4.1.1 串的定义	75
4.1.2 串的基本运算	76
4.1.3 一个有趣的问题	78
4.1.4 串的定长顺序存储	79
4.1.5 模式匹配	81
4.1.6 串的链式存储结构	86
4.1.7 串的应用	87
4.2 广义表.....	88
4.2.1 广义表的定义	88
4.2.2 广义表的存储	89
4.3 数组.....	92

4.3.1 数组的定义和存储	92
4.3.2 特殊矩阵的压缩存储	93
习题 4	96
第 5 章 树	99
5.1 树的概念及基本运算	99
5.1.1 树的概念	99
5.1.2 树的基本术语	100
5.1.3 树的基本运算	101
5.1.4 一个有趣的问题	101
5.1.5 树的存储	102
5.2 二叉树的概念与性质	106
5.2.1 二叉树的概念及基本运算	106
5.2.2 二叉树的性质	107
5.2.3 二叉树的存储	109
5.3 二叉树的遍历	111
5.4 二叉树遍历算法的应用	114
5.5 线索二叉树	116
5.6 树和二叉树	123
5.6.1 树与二叉树的转换	123
5.6.2 二叉树与森林的转换	126
5.7 哈夫曼树及其应用	127
5.8 树的应用	131
习题 5	133
第 6 章 图	135
6.1 图的概念及运算	135
6.1.1 图的概念	135
6.1.2 图的基本运算	138
6.1.3 一个有趣的问题	139
6.2 图的存储	139
6.2.1 数组表示	140
6.2.2 邻接表表示	142
6.3 图的遍历	145
6.3.1 深度优先搜索遍历	145
6.3.2 广度优先搜索遍历	148
6.4 图的连通性问题	151
6.4.1 无向图的连通性	151
6.4.2 最小生成树	152

6.4.3	Prim 算法	153
6.4.4	Kruskal 算法	156
6.5	最短路径	158
6.5.1	单源点最短路径	159
6.5.2	任意一对顶点之间的最短路径	162
6.6	有向无环图的应用	163
6.6.1	AOV 网	164
6.6.2	拓扑排序	165
6.6.3	AOE 网	167
6.6.4	关键路径	167
6.7	图的应用	172
	习题 6	173
第 7 章 排序		176
7.1	排序的基本概念	176
7.2	一个有趣的问题	177
7.3	插入排序	178
7.3.1	直接插入排序	178
7.3.2	折半插入排序	180
7.3.3	希尔排序	182
7.4	交换排序	185
7.4.1	冒泡排序	185
7.4.2	快速排序	187
7.5	选择排序	190
7.5.1	直接选择排序	190
7.5.2	树形选择排序	192
7.5.3	堆排序	193
7.6	归并排序	198
7.7	排序的应用	200
7.8	各种排序方法的综合比较	201
	习题 7	202
第 8 章 查找		205
8.1	查找的基本概念	205
8.2	一个有趣的问题	206
8.3	静态查找表	209
8.3.1	顺序查找法	209
8.3.2	折半查找法	211
8.3.3	分块查找法	214

8.4 动态查找表	216
8.5 B-树	225
8.6 哈希表	233
8.6.1 哈希法与哈希表	233
8.6.2 冲突处理的方法	235
8.6.3 哈希函数的构造方法	238
8.6.4 哈希表的查找	241
8.7 查找的应用	244
习题 8	245
参考文献	249

第1章

绪论

自1946年世界上第一台电子计算机^① ENIAC诞生开始,特别是近20年来计算机产业的飞速发展与应用已远远超出人们对它的预料。摩尔(Moore)定律^②指出:微处理器的速度每过18个月就会翻一番,而半导体芯片的价格每隔18个月就会降低一半。迈特考夫(Metcalf)定律^③指出:网络的价值与上网人数的平方成正比。计算机在处理速度、存储容量、网络化,以及软件的精巧化方面经过数十年的发展,已经渗入到科学、商业和文化等领域中,而智能工程和大数据技术又将令其从量变转向质的飞跃。在科学领域,计算机可以模拟气候变化、破解人类基因;在商业领域,低成本的计算、因特网和数字通信正在改变全球经济;在文化领域,计算类产品,如 iPod、YouTube 和计算机动画等无处不在。

随着人类社会的发展及科学技术在一些关键领域(如生物技术等领域)的突破,不仅产生了大量的数据,而且更新速度很快。这些数据不仅数量巨大、数据类别多,而且数据的维数很高。截至2011年年底,中国气象科学数据共享平台的数据量已达到116TB(1TB=1000GB),可在线共享服务的数据量超过了50TB;截至2015年年底,中国网页数量已达到了2000亿;根据麦肯锡全球研究院(MGI)估计,全球企业2010年在硬盘上存储了超过7EB(1EB=10亿GB)的新数据,而消费者在PC和笔记本等设备上存储了超过6EB新数据;而IDC预计到2020年全球数据量将达到35ZB(1ZB=1000EB)。

2012年3月,美国公布了“大数据研发计划”,该计划旨在提高和改进人们从海量和复杂的数据中获取知识的能力,进而加速美国在科学与工程领域研发的步伐,增强国家安全。根据该计划,美国国家科学基金会、国立卫生研究院、国防部、能源部、国防部高级研究计划局、地质勘探局6个联邦部门和机构共同提高收集、储存、保留、管理、分析和共享海量数据所需的核心技术,增加大数据技术开发和应用所需人才的供给。欧盟方面也有类似的举措,过去几年欧盟已对科学数据基础设施投资1亿多欧元,并将数据信息化基础设施作为

① 第二次世界大战期间,美国军方为了解决计算大量军用数据的难题,成立了由宾夕法尼亚大学的莫奇利和埃克特领导的研究小组,于1946年2月研制成世界上第一台计算机 ENIAC。我国的第一台计算机 103 机则于1958年8月由中科院计算所研制成功,该计算机的研制只花了两年左右的时间。

② 1965年,作为英特尔公司的创始人之一,戈登·摩尔应邀撰写了一篇名为“让集成电路填满更多元件”(Cramming More Components Onto Integrated Circuits)的文章。文中,摩尔对未来半导体元件工业的发展趋势作出了预测——他指出,单块硅芯片上所集成的晶体管数目大约每年(注:1975年,摩尔将周期修正为“每两年”)增加一倍。这一预言后来成为广为人知的“摩尔定律”。摩尔定律在提出后的近30年与现实发展惊人的一致,预计还可能有效持续到2020年左右。

③ 迈特考夫博士(Dr. Metcalfe)于1973年发明了与传统的计算机网络实现方法完全不同的局域网络——以太网(Ethernet)。

Horizon 2020 计划的优先领域之一。

尽管数据是信息和知识的源泉,但数据并不等于信息和知识,关键在于人类如何从中挖掘它们、理解它们。在生命科学技术领域,理解大量生物学数据所包括的生物学意义已成为后基因组时代极其重要的课题,今日生物学数据的巨大积累和分析处理将促使重大生物学规律的发现。然而,与正在以指数方式增长的生物学数据相比,人类相关知识的增长却十分缓慢。一方面是海量的数据;另一方面是人们在生命技术、农业和环保等方面对新知识的渴求,这些新知识将帮助人们改善其生存环境和提高生活质量。这就构成了一个极大的矛盾,这个矛盾使得能否加速这些海量数据的有效处理成为科学技术发展和人类生活质量进一步提高的瓶颈。这个矛盾的解决需要利用计算机技术。

计算机软件是计算机系统的灵魂,没有软件的计算机系统无法工作。由于计算机软件的核心是算法,而算法实质上是加工数据的过程,所以研究数据结构对提高编程能力和设计高性能的算法至关重要。本章介绍数据结构和算法分析的基本概念及程序设计基础。

1.1 数据结构的基本概念

1.1.1 数据结构的实例

【例 1.1】 学生成绩登记表。

表 1.1 中的每一行可以当作一个结点(或称为记录),它由学号、姓名、性别、三门课程的成绩、总分等数据项组成。在这个表中,第一个记录没有直接前驱,称为开始结点;最后一个记录没有直接后继,称为终端结点。除了第一个记录和最后一个记录以外,其余记录都只有一个直接前驱记录和一个直接后继记录。这种结点之间存在着“一对一”的关系,这种关系被称为“线性”关系。根据数据之间的这种逻辑关系,可以定义按学号查询该同学的成绩或在某个同学的后面插入另一个同学的成绩记录等运算。

表 1.1 学生成绩登记表

学号	姓名	性别	操作系统	计算机网络	高等数学	总分
01	丁一	男	90	73	94	257
02	马二	男	67	75	91	233
03	张三	女	87	91	67	245
04	李四	男	90	89	79	258
05	王五	女	97	85	85	267
06	赵六	男	78	93	91	262
⋮	⋮	⋮	⋮	⋮	⋮	⋮

【例 1.2】 井字棋对弈^①问题。

图 1.1(a)是井字棋对弈过程中的一个格局,任何一方只要使相同的 3 个子连成一条直

^① 计算机博弈(也称为机器博弈)被认为是人工智能领域最具挑战性的研究方向之一,是机器智能、兵棋推演、智能决策系统等人工智能领域的重要基础。2015 年 10 月谷歌的人工智能系统 AlphaGo(阿尔法围棋)以 5:0 的战绩完胜欧洲冠军、职业围棋二段樊麾,这是人类历史上围棋人工智能(AI)第一次在公平比赛中战胜职业选手。

线(也可以是一行、一列或一条对角线)即为胜方。如果下一步由“×”方下,则可以派生出 5 个子格局,如图 1.1(b)所示;随后由“○”方接着下,对于每个子格局又可以派生出 4 个子格局……

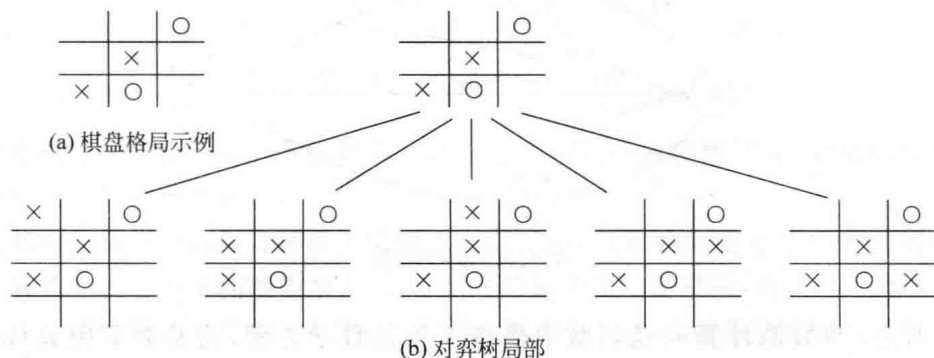


图 1.1 井字棋对弈“树”

若将从对弈开始到结束的过程中所有可能的棋盘格局画在一张图上,即形成一棵倒挂的对弈“树”。“树根”是对弈开始之前的棋盘格局,而所有“叶子”便是可能出现的结局,对弈过程就是从树根沿树枝到达某个叶子的过程。在本例中,对弈中出现的每一个棋盘格局都可以算作一个“结点”。对弈开始之前的棋盘格局没有直接前驱,称为开始结点(根),以后每走一步棋,都有多种应对的棋盘格局,结点之间存在着“一对多”的关系,这种关系被称为“树”关系。根据这种逻辑关系,可以定义求某个棋局有多少个后续棋局,求某个棋局是从哪个棋局发展而来,以及通过对弈树的遍历寻找最佳落子序列等运算。

【例 1.3】七桥问题。

18 世纪,流经哥尼斯堡^①的普雷格河的河湾处有两个小岛,七座桥连接了两岸和小岛。当地流传一个游戏:要求在一次散步中恰好通过每座桥一次,再回到原出发点。设这四块陆地分别为 A、B、C、D,如图 1.2 所示。

1736 年,瑞士数学家欧拉^②(Euler)引入图的概念,证明七桥问题永远无解。欧拉以结点表示四块陆地,以边表示桥,把图 1.2 抽象为图 1.3,即欧拉回路,上述游戏变成图 1.3 中的图形能不能一笔画成的问题。欧拉对七桥问题的结论是:“所有结点的度(一个结点拥有的边数称为度)均为偶数时,原问题才有解”。因此从图 1.3 中可以看出,每一个结点都有多个直接前驱和多个直接后继,也就是说,结点与结点之间存在“多

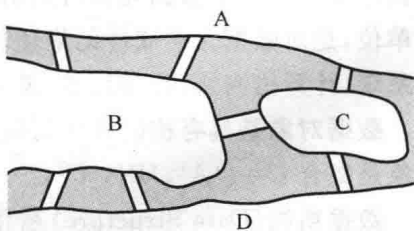


图 1.2 七桥问题

① 今俄罗斯加里宁格勒,是著名的哥尼斯堡七桥问题的发源地。

② 欧拉(Euler, 1707—1783 年),瑞士数学家与自然科学家。1707 年 4 月 15 日出生于瑞士的巴塞尔,1783 年 9 月 18 日于俄国的彼得堡去世。欧拉是 18 世纪数学界最杰出的人物之一,他不仅为数学界做出贡献,还把数学推至几乎整个物理的领域。他是数学史上最多产的数学家,编写了大量关于力学、分析学、几何学、变分法的课本,《无穷小分析引论》《微分学原理》,以及《积分学原理》都成为数学中的经典著作。除了教科书外,欧拉平均以每年 800 页的速度写出创造性论文。1736 年,在经过一年的研究之后,29 岁的欧拉提交了《哥尼斯堡七桥》的论文,圆满解决了这一问题,同时开创了数学的新分支——图论。他去世后,人们整理出他的研究成果多达 74 卷。

对多”关系,这种关系被称为“图”关系。根据这种逻辑关系,可以定义与某一结点相邻的结点有哪些,两个结点间有几条通路等运算。

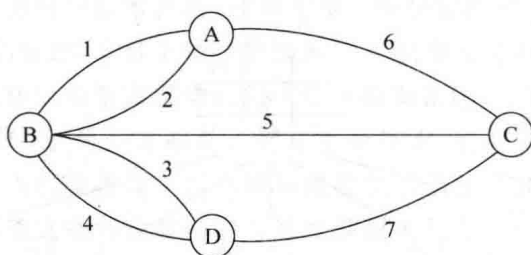


图 1.3 欧拉回路

综上所述,非数值计算问题的数学模型不再是数学方程,而是要研究诸如例 1.1、例 1.2 和例 1.3 所描述的数据之间的这些关系,以及通过数据之间的这些关系可以方便地进行什么运算。因此,简单地讲数据结构是一门研究数据之间的关系及运算的学科。

1.1.2 数据结构的概念

数据是信息的载体,是对客观事物的符号表示。通俗地说,凡是能被计算机识别、存取和加工处理的符号、字符、图形、图像、声音、视频信号、程序等一切信息都可以称为数据。数据可以是数值数据,也可以是非数值数据。数值数据包括整数、实数、浮点数和复数等,主要用于科学计算和商务处理等;非数值数据则包括文字、符号、图形、图像、动画、语音、视频信号等。随着多媒体技术的飞速发展,需要计算机处理的非数值数据已越来越多。

数据元素(记录)是对现实世界中某独立个体的数据描述,是数据的基本单位。在计算机中,数据元素常作为一个整体来处理。一个数据元素可以由若干个数据项组成,在 C 语言程序设计中一个数据元素可以由一个结构体变量表示。数据项是具有独立意义的最小数据单位,是对数据元素属性的描述。例如,在表 1.1 中,每个数据元素由学号、姓名、性别、操作系统、计算机网络、高等数学、总分等数据项组成。

数据对象是具有相同性质的数据元素的集合,是数据的一个子集。例如,字母字符数据对象是集合 $C = \{'A', 'B', 'C', \dots, 'Z'\}$ 。

数据结构(Data Structure)是相互之间存在一种或多种特定关系的数据元素的集合。根据数据元素之间关系的特性。数据结构可由一个二元组 (D, S) 定义,其中 D 是数据元素的有限集, S 是 D 中元素的关系的有限集。通常有以下 3 种基本的数据结构。

- (1) 线性结构:数据元素之间存在“一对一”的关系,如例 1.1。
- (2) 树形结构:数据元素之间存在“一对多”的关系,如例 1.2。
- (3) 图结构或网结构:数据元素之间存在“多对多”的关系,如例 1.3。

图 1.4 描述了这 3 种数据结构。上面这 3 种数据结构可以划分为两大类:线性结构和非线性结构。线性结构的逻辑特征是有且仅有一个开始结点和一个终端结点,并且所有的结点都最多只有一个直接前驱和一个直接后继。非线性结构的逻辑特征是一个结点可能有多个直接前驱或直接后继。非线性结构包括树形结构和图结构。